Supporting Information S3 for 'Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity'

Wilson J. Wright, Kathryn M. Irvine, Emily S. Almberg, and Andrea R. Litt

Model extension if detections are unable to be confirmed

When using a second, error-free method to confirm automated classifications, some call recordings may remain ambiguous. With bat acoustic data, for instance, experts are not able to manually confirm the species for all call files. We still assume that when call files are identified manually that there are no identification errors in this process, but also need to account for the fact that a portion of call files will remain ambiguous for confirmed visits. We illustrate how this additional aspect of the manual confirmation process may be accounted for in the model. As with the other approaches for informing the species classification probabilities, there is an assumption that the call files capable of being confirmed are representative of all calls. In other words, we must still be able to estimate the species classification probabilities using only information from the identifications of call files that were able to be manually confirmed.

To account for this aspect of manually confirming calls, we add an additional parameter for each species, ϕ_k, representing the probability that a call is capable of being manually confirmed. For each confirmed visit, let the number of calls that were manually confirmed to species k and had an automated identification of species k' be represented as $\tilde{C}_{ijkk'}$ for visit j at site i. Based on the properties described previously,

$$[\tilde{C}_{ijkk'} \mid Z_{ik} = 1] \sim \text{Poisson}(\lambda_{ijk}\theta_{k'k}\phi_k)$$

and these counts are directly observable. At the confirmed visits there is also a correspond-

ing set of calls that were not capable of being manually confirmed, but these are latent because the true species identifications are still unknown. These each also follow a Poisson distribution but with rate parameters $\lambda_{ijk}\theta_{kk'}(1-\phi_k)$. For the calls without manual identifications, as with unconfirmed visits, the observed counts are the total with automated identifications to species k'. As defined previously, let $\tilde{C}_{ij\cdot k'}$ represent the unconfirmed calls with automated identifications to species k' for visit j and site i. This count also follows a Poisson distribution, that is

$$\tilde{C}_{ij\cdot k'} \sim \text{Poisson}\left(\sum_{k=1}^{K} z_{ik}\lambda_{ijk}\theta_{kk'}(1-\phi_k)\right),$$

and is directly observable. Incorporating the $(1-\phi_k)$ parameters accounts for the reduction in call rates doing to a portion being manually confirmed for these visits. Again, the key is that the species identification probabilities for the automated process are informed by confirmed detections where possible.